THE CONCEPTION OF STATISTICS AS A TECHNIQUE*

By John Cummings, *Federal Board for Vocational Education*

It seems significant that with few exceptions, those economists who were summoned to Washington during the war period to give counsel and aid in mobilizing our resources for war purposes, worked in the government offices not as economists but as statisticians, and that the services which they organized were officially designated statistical.

Under pressure of the war emergency each newly established war board or council or commission, which engaged actively in mobilization work, proceeded to build up more or less independently of existing services, and at the outset more or less independently of one another, a statistical service of its own. Schedules fluttered out from these offices, and descended in smothering snow-like masses over the whole community, one heavy fall being piled upon another, each federal office functioning as a statistical storm center. In a comparatively short period one office alone sent out over 400 different schedules.

Theories regarding economic conditions and phenomena were for the time being at a discount, and overnight the demand for data became overwhelming. In response to this demand schedule inquiries were instituted regarding stocks of materials, stores of commodities, equipment for war work, number of laborers of different skills, industrial consumption of specific materials, housing conditions, and in general regarding our available material resources.

Government offices, it may be noted, were at the outbreak of the war filled with statistical clerks, statistical technicians, machine operators, expert card punchers, sorters, and counters; and while these technicians with all their paraphernalia were obviously not as a class competent to undertake the statistical enterprises involved in mobilization for war, it is nevertheless probably true that the capacity of the permanent statistical service of the government to undertake the work which had to be done was somewhat underestimated. Some lost motion, confusion, and duplication, some experimenting by those unfamiliar with official statistical practice, and it must be added unfamiliar with available official statistics, might perhaps have been avoided, if the government had proceeded more generally to develop its statistical service from within. Experienced statisticians were, of course, not available either in the government service or in any other service in sufficient numbers to undertake and carry on effectively all

* Read before the eighty-first annual meeting of the American Statistical Association.

of the emergency statistical work which had to be done under pressure, but the policy of expanding existing services might nevertheless, it would seem, have been advantageously adopted more generally than it was, in preference to the policy of building up independent, uncoördinated, temporary services.

One unfortunate consequence of the policy adopted has been the sudden collapse of important statistical enterprises initiated during the war, some of which at least might well have been continued during the reconstruction period, if not permanently. On the whole, the process of statistical demobilization is proving to be nearly as disrupting as was the process of mobilization, and it is being found difficult now to provide for the practical utilization and even for the permanent custodial preservation of statistical war records produced by temporary services which have been or are being disbanded.

But the point of special significance is that practically every inquiry proposed by the government during the war required a statistical answer, and it may be noted incidentally that the science of statistics in fact had its origin in the effort to supply governments with an accurate valuation of resources for waging war. If a statistician may be permitted to say so the simple fact is that economists who engaged in war work seemed to become for the time being statisticians; and economics to become for the time being statistics. Under stress of the war emergency it put away traditional hypotheses, personal philosophies, ethical speculations, and abstract theories, and went out into the hedges and byways to find out things by way of rendering real social service.

Now a quite similar sort of service is just as essential in peace time as in war time. In peace time, as well as in war time, we must mobilize our resources for the achievement of social ends. We must be continuously mobilizing for peace with the end in view to insure in the highest degree our social wellbeing, and in mobilizing for peace, as in mobilizing for war, the inquiries proposed by the government and by agencies of social amelioration are generally inquiries requiring statistical answers.

Under normal peace-time conditions, it may be noted, the work being done by those classified in the government service as economists is, as it has been during the war emergency period, very largely statistical. Official economists are regularly engaged in the conduct of extensive and intensive statistical researches, gathering data by the use of schedules.

Are, then, economists so engaged only undertaking to help out in a field of scientific research which is admittedly short-handed, or is this sort of service strictly economic?

Not infrequently statisticians themselves express the opinion that statistics is in fact not really a science at all. It is, as they conceive it, merely a method, and the statistician is merely a technician who may hire out to do statistical chores in any field indifferently. The economist may of course in any emergency, such as the occurrence of war, do his own statistical chores, but it seems to be assumed that he will perhaps generally prefer to hire them done for him. So of the sociologist and of all other scientific workers, it is assumed that they will each and all on occasion probably want to employ statisticians to do compilation work under direction. In the field of scientific research, it would seem that the statistician cannot operate independently on his own. He must hire out to somebody.

Is then statistics merely a void technique or, if one will, a void formal science on the one hand,—or does the technique or science of statistics embrace a content of data? In the one case the statistician is professionally interested and skilled in the technique and theory of compilation only, and in the other he is professionally interested primarily in the substance, interpretation, and significance of his compilations.

By common usage there would seem to be at least four fairly well differentiated or differentiable concepts as to what statistics are, or what it, if we use the term in the singular, is:

1. Statistics are compiled numerical data of all sorts.
2. Statistics is a method or technique of compiling numerical data of all sorts.
3. Statistics is a formal science of the method or technique of compiling numerical data of all sorts.
4. Statistics is a social science, with a method or technique of its own, a defined field of operation, and a cumulative fund of systematized data or knowledge relating specifically to social conditions and phenomena.

The first of these concepts identifies statistics with figures and tables and that sort of thing indifferently in any field. It is a world-almanac concept. The second and third concepts give to statistics, as a method or technique or science of compilation, a free range in the field of scientific research, but they do not comprehend the data upon which the statistician operates professionally in compiling them into tables. It is not the world-almanac itself, but the method or technique or science of compiling almanacs, that is here proposed as the subject of statistical interest and labor. The fourth concept comprehends the data with the method or technique of compilation, but restricts the range of professionally statistical research to social conditions and phenomena, which is traditionally the special field or preserve of statistical inquiry.

One may perhaps assume that the world-almanac idea of what statistics are will not satisfy the statistician's sophisticated sense as a

definition of his profession, and that he will not write himself down as a mere compiler of the odds and ends of numerical data without rhyme or reason. He will not resign every source of inspiration for work. If he does not pretend to the lofty ideals of a professed science, he will at least pretend to the humble skill of an artisan who has mastered the simple technique of a lowly craft.

Passing on then to the second concept, obviously we cannot say "statistics or statistical data are a method or technique," and when we say "statistics is a method or technique," we clearly oust the data of statistics entirely from our concept, and imply that this mode or art of compilation, separate and distinct from the data compiled, is what we mean by statistics. If we did not mean this we should say, "statistics has—not is—a method or technique."

In this sense statistics is a way or manner or process of doing something. Probably in general what is implied is that it is a way of compiling numerical data, and specifically one way of doing, for example, economics or sociology or politics or biology or astronomy. That is to say, the statistical way of doing each and all of these sciences —one of several ways of doing each of them.

Statistics and statisticians are conceived to be in the service particularly of the social sciences. Statisticians may of course render service also in the biological and the physical sciences, but primarily and properly they ply their skilled trade of compiling data in the social sciences, because these sciences more than any other are dependent upon statistical compilations for the establishment of their basic hypotheses, and for the final verification of their speculative conclusions—or it may be in specific cases, as it has been frequently, for the disestablishment of fictitious hypotheses, and the demolition of false conclusions. In this humble capacity the statistician under close supervision fetches ballast of data for the social sciences, which are naturally speculative and naturally tend to rise into the ethereal regions where the danger of sudden burstings and rapid descents becomes imminent.

Thus the venerable mother of all the social sciences,—for statistics is the mother of the social sciences—is degraded to the position of a menial servant.

Mere compilation of data is in itself, when divorced from the function of interpretation, a very elementary sort of mathematics,—mostly addition and subtraction,—and the technique of compilation, also, is simple and easily mastered. Far less mathematics and technique is required for skilled compilation of data than is required, for example, by

the machinist in many branches of his trade. If the statistician's job is compilation of data of any sort for any one, he is not a very high grade or very skilled artisan, and he clearly cannot rank in the fellowship of economists or politicists or sociologists, except as a sort of handy-boy in the splendid halls and palaces of the social sciences.

This conception of the statistician as a technician is quite consistent, it must be conceded, not only with the idea of purely technical service throughout the broad field of scientific research, but also with the idea which has been in times past rather prevalent in government offices engaged in statistical work. Not uncommonly those administering these offices have felt that a college education, for example, was quite superfluous for a statistician. As they conceived him the statistician was a technician, expert in working out tabulation fields for punching cards, expert in manipulating various types of punching, sorting, tabulating, and calculating machines, expert in reading his own tables into text, and in filling in blanks in standardized texts, preserved inviolate from decade to decade, or from year to year.

For this sort of work, essential as it is, a college education or any very protracted apprenticeship does indeed seem superfluous.

If now the statistician rejects this humble rôle of skilled artisan working under direction, if he pretends to something more than mere expertness as a technician,—if, in a word, he pretends to a separate and distinct science, is the subject matter of that science the method, technique, and mathematics of compilation, including of course, the mathematical determination of correlations, margins of error, interpolations, probable deviations, dispersions, and miscellaneous averages, indices, and rates?

There is here at least something that has altogether the aspect of being scientific, and any one who reads our own and foreign statistical journals, and the various systematic treatises dealing with the mathematical processes of statistical work, will concede that the statistician must, if he is to understand some of the literature of his science, be a fairly good mathematician. Unquestionably he must be able to use, if not himself to develop, mathematical formulas to the extent that may be required in the particular line of statistical work in which he is engaged, and this may mean very considerable expertness as a mathematician.

Mathematics is a universal language in which science expresses itself wherever this language is required to give precision to scientific statements, and the statistician uses this language as he has occasion to use it, just as every other scientific worker uses it, that is to say, as a language for giving precise expression to the conclusions of statistical

research. But in and of itself the purely formal science of statistical compilation and interpretation is applied mathematics, and those who devote themselves entirely to it are essentially mathematicians rather than statisticians.

Practical statisticians know, moreover, that however interesting the mathematical manipulation of data may be in itself as mathematics, data in the raw mass as they come to the statistician are not generally susceptible of any considerable amount of mathematical refinement. Over-refinement of data is professionally recognized as a statistical vice, since it implies accuracy and simplicity in data to a degree which is unwarranted.

Defined as a formal science without specific content of data, statistics seems in fact to assume by definition a rather fictitious character. In the first place, the statistician is made to profess a science which does not comprehend the substance and data of his compilations, and, it may even be said, of his natural professional interests. Secondly, this science of compilation is not really his but the mathematician's, and in it the mathematician is professionally much more expert than the statistician himself. Finally, in the common mine-run of practical statistical work, this science of compilation even in very important lines of statistical research is in itself a poor thing, mostly addition and subtraction, and is not, except in some of its rather ornamental and attenuated refinements as a differentiated branch of mathematics, of a character which justifies its designation as a separate science.

As a professed master of this quite diminutive and ignoble science, it is true again that the statistician may sell his services in the open market to do compilation work in any field, but the chances are that his pretensions to a defined science will not be generally recognized as valid. As a compiler of data for the biologist, he will be for all his professions accepted as merely an artisan, practicing the simple technique of his trade in the field of biology. Probably, if the biological data require any considerable amount of mathematical refinement, the biologist will employ not a statistician but a mathematician. And he will be quite right in doing so.

It is to be noted that the statistician entering any foreign field,—that is to say, any field outside the range of social phenomena,—can carry with him only a very simple technique and rather elementary rules of procedure. He may take with him also the paraphernalia of his trade, card punching, sorting, tabulating devices, adding machines, slide rules, and calculating machines. This he does in his character as an artisan working with a more or less elaborate kit of tools and some degree of technical skill.

It will be freely admitted by statisticians that occasionally the statistician himself, for the precise interpretation of his data, must seek aid of the mathematician, who understands better than the statistician commonly does the harmonics and intricacies of mathematical calculations. And clearly the statistician who does seek such help must be presumed to be interested professionally in the substance and data of his compilations. Otherwise he will be in the position of employing a mathematician to render precisely that service which he himself professes to render, and his pretense to a science will become perfectly transparent and hypocritical.

What then is the statistician professionally interested in? The answer is that his interests lie clearly in the field of social phenomena. In this field statistics as a science originated in the service of princes and principalities; in this field it has rendered continuous service during the past centuries; and in this field it is today rendering service, extending the scope of its researches to embrace every phase of social life, direction, and conduct. This is the traditional field occupied by statistics down through the ages from its remotest origins as an administrative political science, the science of things worth knowing about the state, and developed in subsequent periods as the science of the *Goettliche ordnung* of human affairs, the *Physique Sociale* of Quetelet, the *Political Arithmetic* of Petty, and, finally, as the rounded out science of quantitative valuation of social phenomena.

The term science defines itself. It is knowledge, or as the dictionary puts it, "knowledge of principles or facts; accumulated and accepted knowledge which has been systematized and formulated with reference to the discovery of general truths or the operation of general laws; knowledge classified and made available in work, life, or the search for truth; comprehensive, profound, or philosophical knowledge."

Could any statistician devise a definition which would more certainly embrace in all of its aspects the accumulated and accumulating fund of compiled statistical data relating to social phenomena? Here certainly is a fund of "accumulated and accepted knowledge which has been systematized and formulated with reference to the discovery of general truths or the operation of general laws; knowledge classified, and made available in work, life, or the search for truth"—and the statistician may well emphasize in this definition the word "life." Statistics is knowledge systematized and made available for *life*, not primarily as raw material for the development of some other science, social or natural.

One essential characteristic of any true science is such a fund of

established, universally accepted truths, representing in the mass the actual achievement to date of scientific workers in the given field. Statistics as a social science has this prime characteristic of a true science in a preëminent degree, far surpassing every other social science in this respect of an accumulating fund of systematized knowledge.

"A science," according to Jevons, "teaches us to know, and an art to do, and all the more perfect sciences lead to the creation of corresponding useful arts." Does statistics lead to any useful art or arts?

There is, of course, an art of compiling data, and this art may be conceived to have its origin in the science of statistics; but the art of compiling data is one of the most insignificant of the arts to which statistics leads.

Perhaps the supreme art mothered by statistics is the art of government. One must, however, add to this supreme statistical art all the specialized arts of social conduct, and amelioration—such as those occupied with philanthropy, social hygiene, and education. Not that statistics is the only science back of these several arts of social direction, conduct, and amelioration, but it is traditionally and essentially the primal science of these arts.

Legislation, philanthropy, and every program of social amelioration must take account of actual conditions and tendencies as collective aggregates, and statistics must provide the essential data for every such program, because the collective mass tendencies in social life cannot be uncovered except as they are given statistical measurement, statement, and valuation. Statistics as a social science stands this test of fostering useful arts also excellently well.

Statisticians accept freely the idea of service, and of service especially in the field of the social sciences. They must insist, however, that they render this service not as servants but as masters,—not as either servants or masters of economists or sociologists, but as masters of their own independent procedure and selection in the field of social research. The economist may properly present certain of his problems to the statistician for solution; as may also the statesman, or the expert in social hygiene, or the philanthropist. The prohibitionist may submit to the statistician the problem of determining certain of the social effects of drinking and of prohibition, and the Life Extension Institute may submit its problems. But the ultimate source of the statistician's problems is not economic science, or sociology, or the league to extend prohibition, or an institute to extend life itself, or any program of social amelioration. The ultimate source of the statisti-

cian's problems is the living, growing, changing, complex thing which we call society.

Historically, economics, politics, and sociology have been defined within the field of statistical research as originally conceived. Social sciences by virtue of the fact that they are social rather than natural, deal with aggregates of units and collective or composite dynamic tendencies or forces. They deal with mass effects,—with classes rather than with units or individuals, and every statement regarding a collective aggregate assumes compiled data.

Statistics presents compiled data relating to the production, distribution, exchange, and accumulation of wealth; it relates wealth accumulation to population increase, wealth production to rates of wages and other incomes, volume of credit to price fluctuations. It determines rates of criminality, insanity, and defectiveness for different classes, and thus measures the vital powers of resistance to degeneration and disintegration,—resistance, that is to say, of the individual, regarded as a unit in a collective aggregate, or of the institution of the family, or of some other social institution, in the selective struggle for survival. It measures the breeding power of the race, or stock, or class, against the mortality rate in such detail as may be significant, developing accurate and concise statements of tendencies, out of comparisons of present with past conditions. It measures accurately, and determines the direction of those obscure, insidious, hidden, and involved social tendencies which in the mass determine social evolution or devolution. With these data in hand the economist or the sociologist has a sure foundation of ascertained and demonstrated fact upon which to establish his speculative philosophy.

The statistician cannot perform a similar service for the natural or physical sciences, because these sciences do not deal with collective aggregates except incidentally. In chemistry or physics, and in fact throughout the range of the natural sciences, a single demonstration is frequently sufficient and final. Social laws and tendencies do not present to the scientific investigator that immutability, simplicity, persistence, and finality, which characterizes natural and biological laws. In the field of social phenomena demonstrations must be continuous or periodical. They can never be regarded as final, and all the terms of the social sciences are collective, that is to say statistical aggregates.

In dealing with social phenomena and conditions the statistician must be a free agent. In dealing, for example, with economic phenomena he is not working as an economist, utilizing the statistical

method to develop economic science. He is regarding economic phenomena in all their complexity as presenting a field for independent statistical research. In this field he finds a fund of compiled statistical data relating to past conditions and tendencies, and each year new data are compiled and related to the data brought forward from past periods. The precise measurement of economic tendencies is thus extended, and every calculation is reworked to comprehend a new period.

Similarly throughout the field of social phenomena, the statistician works independently, and unbiassed by any speculative philosophy. Any given statistical induction may be of interest in one of its aspects to the economist, but it may at the same time in other of its aspects be of interest also to others—to the sociologist, the politician, the statesman, the lawyer, the doctor, the merchant, the philanthropist, to organizations of labor or of employers, and to various advocates of social reform and amelioration.

Wherever quantitative valuation is of significance in dealing with social aggregates, the statistician finds his proper employment. Statistics comprehends in all of its varied aspects the precise quantitative measurement, valuation and analysis of social life, and the accumulated results of such measurement, valuation, and analysis continuously carried on from day to day, from year to year, from decade to decade, and from century to century. Every new valuation acquires significance only in proportion as it is related to every similar valuation in the past, and in general with every other quantitative valuation of the past or present with which it may fairly be correlated. Statistics deals with absolute numbers primarily, but always with the end in view of determining tendencies and relativities.

In all of this work the prime characteristic of the statistician is that he has no hypothesis to prove nor any speculative philosophy to develop or validate. He is professionally indifferent to every sort of social propaganda, creed, doctrine, policy, theory, or opinion, and devoted singly to the development of statistical laws expressed as rates, incidences, distributions, averages, frequencies, and correlations, as determined by scientific compilation, refinement, and interpretation of data. By this process the statistician is enabled to forecast the future as regards social trends.

Properly and in character the statistician undertakes the interpretation of his data,—not a partial, but a complete interpretation. He may not be estopped from making an induction by the fact that the induction may have economic or sociological or political significance. All interpretations, even the most elementary, are bound to have

some such significance, and certainly some interpretation of his data must be conceded to the statistician. That an induction happens to be significant is in fact a most excellent reason for making it. The statistician cannot fairly be restricted to the insignificant, simple, and obvious first differences,—to the barren, unrelated first results of his additions and subtractions.

Complete statistical interpretation of data does not, however, comprehend any speculative, ethical, or philosophical elements, since statistical data are in themselves purely quantitative statements, and do not carry any philosophical or ethical or speculative content. There is no "ought" or personality to statistical induction or analysis. The statistician does not pretend to incorporate within his science philosophies not inherent in it but imposed upon it, but he may fairly insist that statistical inductions themselves to their uttermost completeness, extending into every field of social research, are the natural organic developments of statistical science. Statistics does not embrace socialism, because the socialist happens to use some statistics as a basis for developing his peculiar philosophy, nor on the other hand are the statistical inductions which happen to be of interest to the socialist any the less purely statistical on that account. While statistical inductions are philosophically neutral, being in themselves neither conservative nor radical, neither individualistic nor socialistic, they must be nevertheless precisely and completely determined, and made available as the data of all real social science and of all social philosophy.

Purely hypothetical systems of reasoning may very well serve a real scientific end in a primitive stage of scientific development. Such systems have undoubtedly, for example, facilitated the primitive development of economics, but in proportion as data become available the economist is bound to undertake something more than the development of flawless trains of reasoning, originating in purely hypothetical assumptions and terminating therefore in conclusions of doubtful real significance. He is bound rather to undertake to explain economic phenomena in all their baffling complexity, to develop every conclusion from data, and to eliminate everything in the nature of a hypothesis.

So of other social sciences, it is true that they must all gradually abandon primitive speculations in the process of substituting data for hypotheses. It is admittedly much more difficult to develop a living science-than it is to perfect a dead system of logic, but inevitably a social science which remains hypothetical, however flawless and perfect it may be, tends to become an inert, closed, rigid, uninspired system.

In the broad field of social research, it may be noted the vital prob-

lems have generally been precisely those implied in the traditional "ifs," which have been passed on from generation to generation,—such for example as the very large "if" implied in Malthusianism. Statistics would substitute data for these "ifs," not data in the raw, but data compiled, analyzed, interpreted. For every "if" it would substitute the final and ultimate statistical induction—not any philosophy, or doctrine, or ethics, or speculation imposed upon the data, but the very quintessence of the data, the pure, statistical, socially significant distillation of the data.

Every science has undergone or is undergoing this sort of sea-change from something strange into something real. The age of astrology and alchemy passes for all sciences, although no science easily frees itself entirely from its halo of hypotheses and mysticism. Psychology has developed out of painstaking and persistent experimentation, observation, and self-analysis, but it is still more or less involved in mystical speculations regarding the super-ego, the power of telepathy, our astral bodies, spiritism, levitations, and other strange phenomena which the science cannot comprehend or even concede to be real. So in other lines of scientific research, even today as has been true in the past, at a given point speculation supersedes science. In the social sciences, as in other sciences, speculation tends to supersede science at some point, and it must be added that at this point dissention and contention begin, because there is beyond the confines of science no sure foundation for reasoning or philosophy or ethics.

It is perhaps worth noting that stupendous errors have been committed in the development of social policies, philosophies, and speculations by the disassociation of scientific conclusions from the data upon which those conclusions have been based, by the specious substitution of social for natural or particular terms, or by the assumption that conclusions hypothetically or historically true are true absolutely for all time. War, for example, is proposed as a selective struggle and militarism thus becomes a natural political order. Again, because some forms of government are or have been evil, and the exercise of some powers by any form of government is or may be oppressive, the doctrine of *laissez faire* is developed, and government itself is declared to be essentially evil. Or again, poverty is declared to be a natural phenomenon because "if" population tends to increase faster than the food supply poverty is clearly inevitable. From period to period one or another falsely grounded or superannuated social philosophy has exerted its hypnotic spell over whole communities, seeming to tolerate, turn and turn about, now this and now that sort of social evil, and even

to impose as a social duty the commission of certain outrages which a later age frankly condemns. In the field of social phenomena the hypnotic spell of false generalizations developed into social philosophies, policies, and ethics, has on occasion even tended to paralyze and render impotent organized agencies of social betterment.

Science itself does not commit blunders of this sort, but they are committed in the name of science, and particularly of primitive social science.

For these hypotheses and speculations especially, which have been or may be the occasion of blundering and misleading and contention in the field of social phenomena and conduct, statistics would substitute statistical inductions. It is the statistician's function to measure social tendencies precisely, not to reason or philosophize about hypothetical tendencies. It is his function to determine precisely rates and frequencies and correlations and probabilities in so far as these can be uncovered in collective social aggregates. This mass of living data which comprehends the future with the present and the past, this quantitative summing-up and statement of social phenomena and of social values constitutes the science of statistics. Skill and technique are means to an end in statistical work as in every branch of science and art. They are not in themselves ends, and will not serve to guide the statistician in electing the right course. He must bring to his work all the resources and guidance of a rounded-out science, which sums up past achievements, prescribes its peculiar technique, and gives direction to present endeavor.